

REMARKS

In the patent application, claims 1 and 3-48 are pending. In the final office action, all pending claims are rejected.

Applicant has canceled claims 42-48, amended claims 1, 3-25 and 27-41, and added new claims 49-56.

Claim 1 has been amended to include the limitations of obtaining, for each of a plurality of consecutive time intervals, one or more parameters from an audio signal, said one or more parameters relating to audio characteristics of the audio signal, and segmenting the audio signal into a plurality of segments based on the parameters obtained for the consecutive time intervals.

The support for the amendment can be found in Figure 4; p.11, lines 26-31; p.13, lines 8-11, lines 21-24.

Claims 19 and 22 have been amended to include the limitations that the audio data indicative of a plurality of segments of an audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal, and wherein the plurality of segments are obtained based on the parameters extracted for the consecutive time intervals, and the audio data is indicative of the parameters in an adjusted representation.

The support for the amendment can be found in Figure 4; p.11, lines 26-31; p.13, lines 8-11, lines 21-28.

Claim 27 has been amended to include the limitation that that the audio data indicative of a plurality of segments of an audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal, and wherein the plurality of segments are obtained based on the parameters extracted for the consecutive time intervals, and the audio data is indicative of the parameters in an adjusted representation.

The support for the amendment can be found in Figure 4; p.11, lines 26-31; p.13, lines 8-11, lines 21-28.

Claim 31 has been amended to include the limitation that the audio data is indicative of a plurality of segments of an input audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating

to audio characteristics of the audio signal, and wherein the plurality of segments are obtained based on the parameters extracted for the consecutive time intervals and encoded with a plurality of encoding settings based on the audio characteristics, the audio data indicative of the parameters in an adjusted representation.

The support for the amendment can be found in Figure 4; p.11, lines 26-31; p.13, lines 8-11, lines 21-28.

Claim 32 has been amended to be dependent from claim 19.

Claims 3-18, 20, 21, 23-25, 28-30 and 33-41 have been amended to change the wording.

New claims 49, 51 and 53 are dependent from claims 1, 19 and 22 and include the limitation that the parameters are obtained from the audio signals in regular time intervals.

The support for these new claims can be found on p.11, lines 26-31; p.14, lines 19-20; p.15, lines 8-12.

New claims 50, 52 and 54-56 are dependent from claims 1, 19, 22, 27 and 31 and include the limitation that the segmenting or the plurality of segments are based on the similarity in the parameters among consecutive time intervals.

The support for new claims 50, 52 and 54-56 can be found on p.18, lines 17-19.

No new matter has been introduced.

At section 2 of the final office action, claims 1, 3-48 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. The Examiner states that claims 1, 19, 22, 27, 31 and 32 have the limitation of segmenting audio signals based upon audio characteristics, but it is not clear as to which segmenting aspect of the disclosure this refers. The Examiner states that the specification discloses two aspects of segmenting:

1) a typical audio encoder that extracts audio signal information (outputting segments based upon voice/unvoiced, silence decision denoted as line 110 into the sub-block 12 in Figure 4, generating segmented audio with associated parameters 112 (p.13, lines 8-14 of the specification); and

2) the sub-block 20 re-segments sequence of initial segments based on degree of voicing, etc., derived from speech parameters (Figure 4, p.15, lines 1-17 of the specification).

The Examiner further states that the current claim scope does not distinguish between these two sections of the applicant's disclosure and as such, these claims are rejected under 35 U.S.C. 112, second paragraph. For art related examination purposes only, the Examiner will interpret the claim scope to read upon the first section (aspect) discussed above, namely, the encoder of Figure 4 that encompasses only line 110, sub-block 12, and line 112. The dependent claims do not remedy the deficiencies of the independent claims, and as such, are also rejected under 35 U.S.C. 112, second paragraph.

It is respectfully submitted that the present invention is directed to a method and device for enhancing the coding efficiency of a parametric speech coder (p.11, lines 18-20). In a typical parametric speech coder, the parameters extracted at regular intervals include linear prediction coefficients, speech energy, pitch and voicing information (p.11, lines 26-28). Based on the parameters related to speech energy and voicing, a simple segmentation algorithm can be implemented (p.12, lines 1-2). Figure 4 is a block diagram showing the speech coding system, according to the present invention. In Figure 4, block 12 provides parameters 112 to the compression module 20 (p.13, lines 9-11). Based on the behavior of the parameters, the compression module 20 carries out a number of steps, including segmentation of the input speech signal and efficient quantization of the derived parameters (p.13, lines 21-28).

A typical speech coder, such as block 12, receives an input speech signal 110 and outputs parameters 112. Typically, such a speech coder is configured to carry out the following steps:

- 1) receiving an input speech signal;
- 2) sampling the input speech signal at consecutive time intervals – in general, the consecutive time intervals include voiced (v), unvoiced (u) and silent (s) samples;
- 3) transforming v, u and s samples into parameters (here referred to as V, U and S) indicative of the audio characteristics of the samples; and
- 4) outputting the parameters V, U and S as audio data.

In plain English, “outputting” means generating, providing or producing, and “segmenting” means dividing. In a parametric coder, segmentation only takes place at step 2 where the input signal is sampled at consecutive time intervals. In particular, the parameters extracted at regular intervals in a parametric encoder. Thus, the sampling of the input speech

signal into samples at consecutive time intervals is not based on the audio characteristics of the samples.

Let us assume that the input speech signal 110 is sampled into a sequence of samples such as: uuussvvvssuuvvvv. After the transformation step, the parameters 112 in the outputted audio data will be UUUUSSVVVSSUUVVVV. The parameter extraction unit 12 outputs whatever it is given to transform and, therefore, the sequence of the parameters in the outputted audio data corresponds to the sequence of samples. In the outputted audio data, the length of the unvoiced segment UUUU is the same as the length of the consecutive voiced sample uuuu. Although the outputted parameters 112 include unvoiced segments, silent segments and voiced segments, the “outputting” function does not “divide” the audio data based on whether the parameters are U, V or S. After the transformation from u to U, from v to V and from s to S, there is no algorithm or mechanism in the parametric coder to decide how the audio data is divided before outputting. Since the sampling step that yields the sample sequence uuussvvvssuuvvvv is not based on the audio characteristics of the samples, the outputting step that provides the parameter sequence UUUUSSVVVSSUUVVVV is not based on the audio characteristics of the parameters.

Thus, in Figure 4, block 12 does not involve segmenting the audio signal into a plurality of segments based on the parameters. Segmentation that is based on the parameters 112 takes place in the compression module 20.

For the above reasons, applicant respectfully requests that the 112 rejection be withdrawn.

At section 3, claims 1, 3-14, 19-21, 26-37, 39-44, 46-48 are rejected under 35 U.S.C. 102(b) as being anticipated by *Gersho et al.* (U.S. Patent No. 6,311,154, hereafter referred to as *Gersho*).

At section 5, claims 15-18, 22-25, 38, 45 are rejected under 35 U.S.C. 102(e) as being anticipated by *Sinha et al.* (U.S. Patent No. 7,191,136, hereafter referred to as *Sinha*)

In rejecting claim 1, the Examiner states that *Gersho* teaches segmenting {partitioning or classifying} the audio signal {speech} into a plurality of segments {frames} (partitioning

samples of a speech signal into frames, col. 4, lines 25-27) based on the audio characteristics {classes} of the audio signal (classifying the speech signal in each frame into one of a plurality of classes, col. 4, lines 25-27); and encoding the segments {frames} with different encoding settings {excitation} (encoding an excitation for the frame using one of a plurality of excitation coding...selected according to the class of the frame, col. 4, lines 30-33).

It is respectfully submitted that claim 1 includes the limitations of obtaining, for each of a plurality of consecutive time intervals, one or more parameters from an audio signal, said one or more parameters relating to audio characteristics of the audio signal, segmenting the audio signal into a plurality of segments based on the parameters obtained for the consecutive time intervals; and encoding the segments with different encoding settings.

In col.4, lines 23-34, *Gersho* discloses:

Further in accordance with this invention there is provided a method for coding a speech signal that includes steps of (a) partitioning samples of a speech signal into frames; (b) deriving a residual signal for each frame; (c) classifying the speech signal in each frame into one of a plurality of classes; (d) identifying the location of at least one window in the frame by examining the residual signal for the frame; (e) encoding an excitation for the frame using one of a plurality of excitation coding techniques selected according to the class of the frame; and, for at least one of the classes, (f) confining all or substantially all of non-zero excitation amplitudes to lie within the windows.

In the above passage, *Gersho* discloses a method in which the speech signal containing speech samples is partitioned or segmented into frames in step (a) and the speech signal in each frame is classified into one of the classes in step (c). As shown in Figure 8, *Gersho* uses a high-pass filtering block 30 to filter the input speech into high-pass filtered speech (col.14, lines 56-57). The high-passed filtered speech is divided into non-overlapping “frames” of 160 samples each (col.15, lines 7-8). According to *Gersho*, all the frames contain the same number of

samples. The filtered speech is provided to two separate modules: parameter estimation module 32 and open-loop classifier module 34. The residual signal and the model parameters for each of the frames are extracted or estimated by the model parameter estimation module 32. The classification information OLC(m) is provided to an excitation encoding and speech synthesis module 42. The residual signal and parameters, along with the classification information, are used for speech synthesis.

According to *Gersho*, the high-passed filtered speech is segmented into frames such that each frame has a fixed number of samples, and the estimation of speech parameters and the classification for the frames are carried out after the filtered speech is divided. This indicates that the partitioning of the speech signal is not based on the parameters.

Gersho does not disclose or suggest segmenting the audio signal into a plurality of segments based on the parameters obtained for the consecutive time intervals.

For the above reasons, *Gersho* fails to anticipate claim 1.

In rejecting claims 19 and 27, the Examiner states that *Gersho* teaches an input for receiving audio data indicative of parameters in the adjusted representation (input applied to element 14, Figure 3) and a module responsive to the audio data for generating the audio signal based on the adjusted signals and the characteristics of the audio signal (Figure 3).

It is respectfully submitted that claim 19 is directed to a decoder which comprises:

an input for receiving audio data indicative of a plurality of segments of an audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal, and wherein the plurality of segments are obtained based on the parameters extracted for the consecutive time intervals, and the audio data is indicative of the parameters in an adjusted representation; and

a module, responsive to the audio data, for generating a further audio signal based on the adjusted representation and the encoding settings.

In Figure 3, *Gersho* discloses a speech encoder 12 for obtaining a smoothed energy contour of a speech residual signal (col. 5, lines 38-40). The speech encoder 12 has a fixed

frame structure (basic frame structure) and each basic frame is partitioned into M equal length subframes (col.7, lines 18-27). In conventional AbS coding schemes, the excitation signal for each subframe is selected by a search operation. However, an adequate precise representation of the excitation segment is difficult to obtain (col.7, lines 28-33). *Gersho* discloses a method for identifying the location of the excitation activity into the subframe by examining a smoothed energy contour of the linear prediction residual (col.7, lines 34-50). Figure 3 depicts an encoder wherein a linear prediction (LP) whitening filter 14 for forming the residual signal from the input speech in order to obtain a smooth energy contour and to identify the energy peaks (col. 8, line 53 to col.9, line 2).

In Figure 3, the block 14 is a filter for obtaining a residual signal by filtering an input speech.

Gersho does not disclose a decoder that comprises an input for receiving audio data indicative of a plurality of segments of an audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal, and wherein the plurality of segments are obtained based on the parameters extracted for the consecutive time intervals, and the audio data is indicative of the parameters in an adjusted representation.

For the above reason, *Gersho* fails to anticipate claim 19.

For the same reasons, *Gersho* also fails to anticipate claim 27.

In rejecting claim 31, the Examiner states that *Gersho* discloses implementing a cell phone system (col. 6, lines 33-36).

It is respectfully submitted that claim 31 includes the limitation of an input module for receiving audio data from at least one of the base stations, the audio data indicative of a plurality of segments of an input audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal, and wherein the plurality of segments are obtained based on the parameters extracted for the consecutive time intervals and encoded with a plurality of encoding settings based on the audio characteristics, the audio data indicative of the parameters in an adjusted representation.

As with claims 19 and 27 above, *Gersho* does not disclose an input for receiving audio data indicative of a plurality of segments of an audio signal, wherein one or more parameters are extracted from the audio signal for each of a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal, and wherein the plurality of segments are obtained based on the parameters extracted for the consecutive time intervals, and the audio data is indicative of the parameters in an adjusted representation.

For the above reasons, *Gersho* fails to anticipate claim 31.

In rejecting claim 22, the Examiner states that *Sinha* discloses a method and device for encoding as claimed.

It is respectfully submitted that claim 22 includes the limitations of
an input for receiving audio data indicative of parameters obtained from an audio signal in a plurality of consecutive time intervals, the parameters relating to audio characteristics of the audio signal; and

an adjustment module for adjusting one or more of the parameters for providing an adjusted representation of the parameters, wherein said adjusting comprises segmenting the audio signal into a plurality of segments based on the parameters obtained for the consecutive time intervals and encoding the segments based on one or more of a plurality of encoding settings.

Sinha discloses a method for improving an audio compression scheme, such as perceptual audio coding (PAC). In a conventional PAC scheme, as shown in Figure 1, the input signal is segmented into frames to be stored in a frame buffer 104. The frames are then processed through a long-term predictor 106 and a short-term predictor 108 for linear predictive analysis (col. 2, lines 13-16). Each of the audio frames in PAC consists of 1024 pulse code modulated (PCM) samples (col.3, lines 52-56). According to *Sinha*, as the input speech signal is segmented into frames of 1024 PCM samples, the speech signal is simultaneously provided to a low-pass filter 402 for obtaining compressed information consisting of coded low frequency components, and to a high-pass filter 404 for obtaining a parametric representation of the high frequency components based on a non-linear model 406. The parametric representation is updated every audio frame in order to estimate the non-linear parameters 408 (col.4, lines 43-59).

According to *Sinha*, segmentation of the speech signal into frames takes place before the high-pass and low-pass filtering and before the non-linear parameters 408 are obtained. Furthermore, the parametric representation is updated every audio frame.

Sinha does not disclose or suggest segmenting the audio signal into a plurality of segments based on the parameters obtained for the consecutive time intervals.

For the above reasons, *Sinha* fails to anticipate claim 22.

As for claims 3-18, 20-21, 23-26, 28-30 and 32-41, they are dependent from claims 1, 19, 27 and 31 and include further limitations. For reasons regarding claims 1, 19, 27 and 31 above, *Gersho* also fails to anticipate claims 3-18, 20-21, 23-26, 28-30 and 32-41.

New claims 49-56 are dependent from claims 1, 19, 27 and 31 and include further limitations. For reasons regarding claims 1, 19, 27 and 31, claims 49-56 are distinguishable over *Gersho*.

CONCLUSION

Claims 1, 3-41 and 49-56 are allowable. Early allowance of claims 1, 3-41 and 49-56 is earnestly solicited.

Respectfully submitted,



Kenneth Q. Lao

Attorney for the Applicant
Registration No. 40,061

WARE, FRESSOLA, VAN DER SLUYS
& ADOLPHSON LLP
Bradford Green, Building Five
755 Main Street, P.O. Box 224
Monroe, CT 06468
Telephone: (203) 261-1234
Facsimile: (203) 261-5676
USPTO Customer No. 004955